Impaired motor-to-sensory transformation mediates auditory hallucinations

Fuyin Yang[a,b,c] , Hao Zhu[c,d], Xinyu Fang[a], Lingfang Yu[a], Zenan Wu[a], Siqi Li[b,c], Chen Zhang[a,1], Xing Tian[d,c,b,1]

a Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China
b Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China
c NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, 3663 Zhongshan Road North, Shanghai, 200062, China
d Division of Arts and Sciences, New York University Shanghai, 1555 Century Avenue, Shanghai, 200122, China

Correspondence to:

Xing Tian, Email: xing.tian@nyu.edu

Chen Zhang, Email: zhangchen645@gmail.com

Abstract

Distinguishing reality from hallucinations requires efficient monitoring of agency. It has been hypothesized that a copy of motor signals, termed *efference copy (EC) or corollary discharge (CD*), suppresses sensory responses to yield the sense of agency; impairment of the inhibitory function leads to hallucinations. However, how can the sole absence of inhibition yield positive symptoms of hallucinations? We hypothesize that selective impairments in functionally distinct signals of *CD* and *EC* during motor-to-sensory transformation cause the positive symptoms of hallucinations. In an electroencephalography (EEG) experiment with a delayed articulation paradigm in schizophrenic patients with (AVHs) and without auditory verbal hallucinations (non-AVHs), we found that preparing to speak without knowing the contents (general preparation) did not suppress auditory responses in both patient groups, suggesting the absent of inhibitory function of *CD*. Whereas, preparing to speak a syllable (specific preparation) enhanced the auditory responses to the prepared syllable in non-AVHs, whereas AVHs showed enhancement in responses to unprepared syllables, opposite to the observations in the normal population, suggesting that the enhancement function of *EC* is not precise in AVHs. A computational model with a virtual lesion of an inhibitory inter-neuron and disproportional sensitization of auditory cortices fitted the empirical data and further quantified the distinct impairments in motor-to-sensory transformation in AVHs. These results suggest that 'broken' *CD* plus 'noisy' *EC* causes erroneous monitoring on the imprecise generation of internal auditory representation and yields auditory hallucinations. Specific impairments in functional granularity of motor-to-sensory transformation mediate positivity symptoms of agency abnormality in mental disorders.

Key words: Schizophrenia, Auditory hallucinations, Sensorimotor integration, Agency, Internal forward model, Computational psychiatry

Introduction

Perceptual experiences can be induced by sensory processing (Hickok & Poeppel, 2007; Ungerleider & Haxby, 1994) as well as constructed without external stimuli, such as memory retrieval (Wheeler et al., 2000) and mental imagery (Kosslyn et al., 2001; Li et al., 2020; Tian & Poeppel, 2013; Tian et al., 2016; Zatorre & Halpern, 2005). Such distinct causes of perceptual experiences necessitate efficient monitoring to distinguish the inducing sources; failure of monitoring may result in hallucinations (McGuire et al., 1995; Tian & Poeppel, 2012). For example, patients with auditory hallucinations, the core symptoms of schizophrenia, often 'hear' voices in the absence of sound (Ford et al., 2007). Patients may fail to distinguish between their thoughts (e.g., inner speech, Alderson-Day & Fernyhough, 2015) and externally generated voices, resulting in a reduced ability to recognize thoughts as self-generated (Feinberg, 1978; Frith, 1987). That is, the symptoms of hallucinations have been attributed to the malfunction of self-monitoring (Frith, 2015; Stephan et al., 2009).

Self-monitoring can be achieved with internal forward models (Kawato, 1999; Wolpert & Ghahramani, 2000), in which a copy of motor signals, termed corollary discharge (*CD*) (Sperry, 1950) or efference copy (*EC*) (von Holst & Mittelstaedt, 1950), transmits to sensory regions (motor-to-sensory transformation) and suppresses sensory neural activities. The internal forward models have been evident ubiquitously across the animal kingdom (Crapse & Sommer, 2008) and the sensory suppression has been hypothesized as an index for signaling the impending reafference as the consequences of an agent's own actions -- the sense of agency (Blakemore et al., 1998; Houde et al., 2002; Kilteni & Ehrsson, 2017; Ross et al., 2001). That is, the suppression function of motor signals may provide an automatic computation to distinguish the sensory neural responses that are either internally induced or evoked by external stimulation (Blakemore & Decety, 2001; Blakemore et al., 1998). Impairment of the inhibitory functions in motor-to-sensory transformation may result in a malfunction of self-monitoring and lead to auditory hallucinations (Feinberg & Guazzelli, 1999; Ford et al., 2001; Stephan et al., 2009).

However, how can the monitoring function of agency inhibit the sensory processing, while at the same time constructing the positive symptoms of auditory hallucinations? Hallucinations are perceptual-like experiences that require specific neural representation activated without sensory stimulation (Waters et al., 2012). The sole inhibitory function of the motor copies cannot fully explain the positive symptoms of hallucinations and is challenged by recent empirical findings. For example, action-induced enhancement has been found in a subset of sensory cortices in addition to action-induced suppression (Eliades & Wang, 2005, 2008; Enikolopov et al., 2018; Flinker et al., 2010; Singla et al., 2017). Auditory neural representations are constructed in auditory working memory based on covert speaking (Chu et al., 2022; Li et al., 2020; Tian et al., 2018). This motor-based auditory working memory (i.e. inner speech) that activates specific neural representation may be misattributed to external sources because of the impaired self-monitoring function in schizophrenia (Fletcher & Frith, 2009; Ford & Mathalon, 2004). All these recent results hint that the copy of motor signals may have a function that sensitizes the sensory cortices in addition to the inhibitory function for monitoring agency. The combination of two complementary functions may mediate the positive symptoms of auditory hallucinations.

A recent study provided preliminary evidence supporting a hypothesis of distinct modulatory functions of motor signals on perceptual processes (Li et al., 2020) -- the *CD* function is generic motor discharge available throughout the course of action and can inhibit processes in the connected sensory regions for indicating all possible sensory consequences of actions (Zheng et al., 2022) (Fig. 1A, cyan arrow). Whereas the *EC* function is a copy of a specific motor plan and selectively enhances the sensitivity to sensory reafference targets caused by actions (Fig. 1A, red arrow). Based on this theoretical framework, in this study, we hypothesize that selective impairments in *CD* and *EC* functions mediate the positive symptoms of auditory hallucinations. Specifically, the CD inhibitory function that should be available in the early stage of motor intention does not operate normally in all schizophrenia patients (Fig.1B&C, grey arrows) -- the negative symptoms (e.g. lack of desire) in patients without AVHs may have weaker movement intention and hence diminish CD signals at the beginning

stage of speaking (Fig. 1B, only left part of CD arrow is grey); whereas in patients with AVHs, in addition to the negative symptoms, the deficits in the self-monitoring of agency is manifested in the impairment of the inhibitory function of CD throughout the entire course of action (Fig. 1C, the entire CD arrow is grey). When EC is available after specific movement plans have been formed, the enhancement function of EC on the prepared speech target is intact in patients without AVHs (Fig. 1B, red arrow). Whereas, in patients with AVHs, the EC enhancement function is imprecise (Fig. 1C, hatched red arrow) and modulates both target reafference and its neighboring auditory units (potential causes of various perceptual-like auditory and verbal contents during hallucinations). That is, the positive symptoms of auditory hallucinations are mediated by the combination of a 'broken' CD and 'noisy' EC. This hypothesis of distinct impairments in motor-to-sensory transformation predicts that 1) the suppression effects of general speech preparation (e.g. preparing to speak without knowing what to say) that was observed in normal population would be absent in both patients with and without AVHs, and 2) patients without AVHs would show identical enhancement effects of specific speech preparation (e.g. preparing to speak a given syllable) as in normal population, but the modulation effects of specific speech preparation would be different in patients with AVHs from those in non-AVHs and normal population.

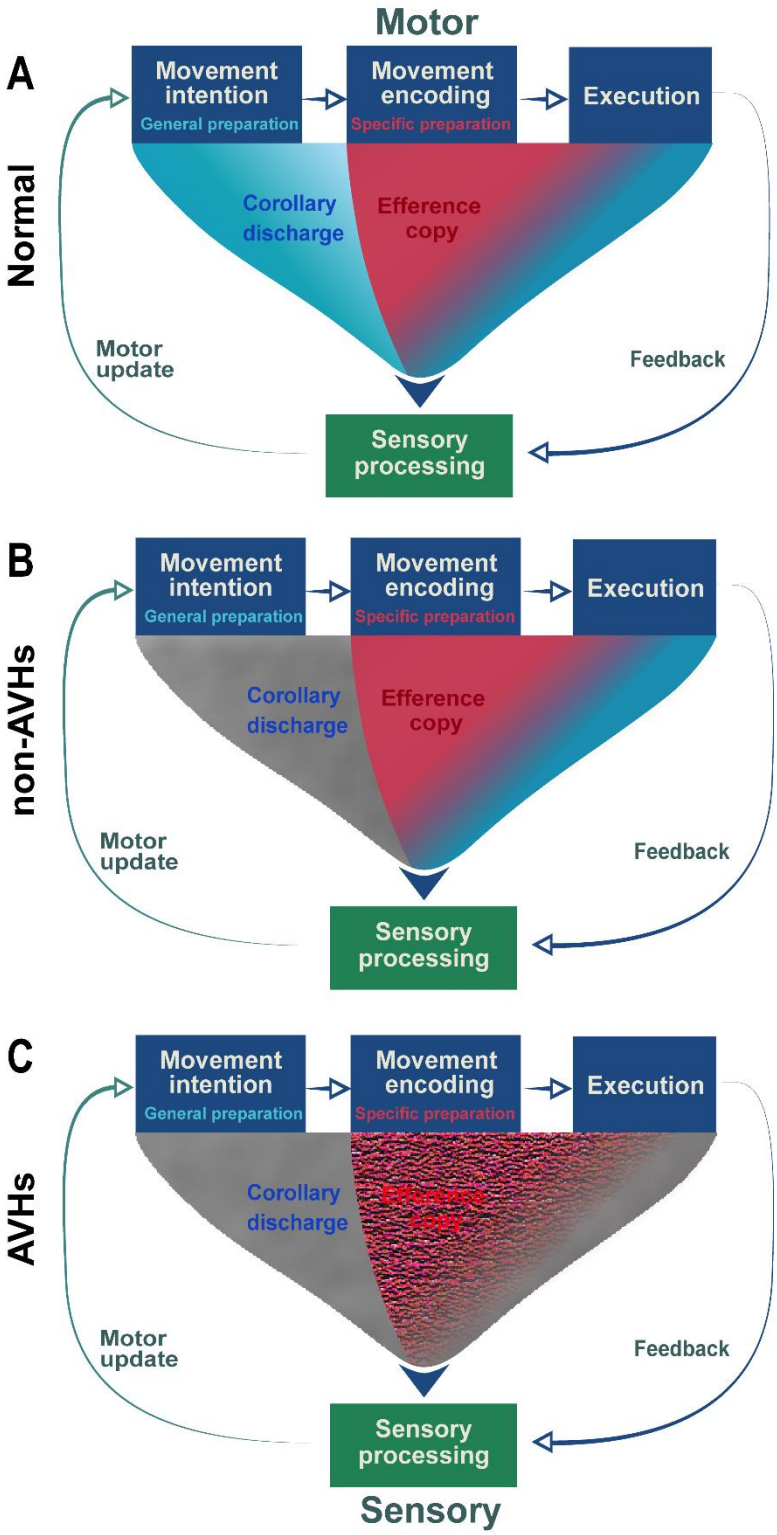

Figure 1. Schematics of distinct functions of motor signals in motor-to-sensory transformation across temporal stages of action in normal and clinical populations. A) The distinct inhibition and enhancement functions in motor-to-sensory transformation in the normal population. Corollary discharge (CD) is a general discharge signal from the motor system that does not necessarily include any content information. CD is available at all stages of motor processes and can onset as early as in motor intention (i.e. general preparation, for example, preparing to speak without knowing what to say). The function of CD is

inhibiting all sensory regions that are connected with the activated motor system, indicating the impending sensory consequences of actions and hence yielding the sense of agency (Fig. 1A, cyan arrow). Efference copy (EC), a duplicate of the planned motor signals, is available during motor encoding (i.e. specific preparation, for example, preparing to speak a specific speech). The copy of detailed action codes selectively boosts the sensitivity of neural responses to the sensory target of actions (Fig. 1A, red arrow). B) The intact enhancement but possible deficits in inhibition functions in motor-to-sensory transformation in schizophrenia patients without auditory hallucinations (non-AVHs). The negative symptoms of non-AVHs patients may cause weak motor intention that leads to diminished inhibitory function of CD at the beginning stage of an action (Fig. 1B, grey in the left part of the CD arrow), whereas the function of CD in the following motor processes (Fig. 1B, blue in the right part of the CD arrow, identical to that in Fig. 1A) as well as the specific modulation effects of enhancement in EC are preserved (Fig. 1B, red arrow, identical to that in Fig. 1A). C). The impaired inhibitory function of CD and imprecise enhancement function of EC in schizophrenia patients with auditory hallucinations (AVHs). The malfunctioned monitoring of agency in auditory hallucinations is mediated by the impaired inhibitory function of CD throughout the course of motor processes (Fig. 1C, entire grey arrow). Moreover, the positive symptoms of auditory hallucinations -- random perceptual-like experiences without corresponding acoustic stimulations – would be mediated by imprecise EC (Fig. 1C, hatched red arrow) that could activate multiple auditory neural representations around the sensory target of the action. That is, the positive symptoms of auditory hallucinations are an emergent property of the impaired sensorimotor systems in which a 'broken' CD misattributes the inducing sources of the auditory neural representations that are activated by a 'noisy' EC without external stimulations. (We are requiring permission for using Fig.1A from the original publisher.)

Result

Demographic and clinical data

Table S1 shows the participants' demographic data and the clinical variables. One-way ANOVA analyses revealed a significant difference among three groups (AVHs, non-AVHs, and normal) in the GP task for age ($F(2,51) = 5.48$, $p = 0.007$), education ($F(2,51) = 8.58$, $p = 0.001$) and in the SP task for age ($F(2,51) = 5.72$, $p = 0.006$), education ($F(2,51) = 4.52$, $p = 0.016$). Fisher's LSD post hoc tests revealed no significant differences between AVHs and non-AVHs groups in age and education in GP and SP tasks (all $p > 0.05$). There was no significant difference in height, and weight among the three groups (all $p > 0.05$). The chi-square test showed no significant differences among the three groups in gender. Further, the positive symptom scores ($t(1,33) = 2.415$, $p = 0.021$) and PANSS total score ($t(1,33) = 2.191$, $p = 0.036$) were significantly higher in the AVHs group than in the non-AVHs group. The negative and general

psychopathology scores were not significantly different in the AVHs and the non-AVHs groups. Neither the age of onset nor the duration was significantly different between the two groups.

Behavioral preparation effects in the delayed articulation task

In the AVHs group (Fig.2C), a repeated-measure one-way ANOVA on RTs found a significant main effect of preparation ($F(3,54) = 85.267$, $p < 0.0001$, partial $\eta^2 = 0.826$). Further analysis revealed that the onset of articulation was consistently faster after preparation. Specifically, RTs of articulation were the fastest after SP (mean = 442.97 ms; SD = 23.66) among all four conditions ($F(3,72) = 24.81$, $p < 0.0001$, partial $\eta^2 = 0.508$). Articulation after GP (mean RT = 603.53 ms; SD = 19.40) was faster than immediate vocalization without preparation (NP, mean RT = 742.09 ms; SD = 34.85) ($t(18) = 5.167$, $p < 0.0001$, $d = 0.883$). RTs were also significantly shorter when participants performed GP without sound probes (GP_{NS}, mean RT = 646.49 ms; SD = 15.53) than NP ($t(18) = 3.96$, $p < 0.001$, $d = 0.791$). These behavioral results suggested that participants engaged in speech preparation. More importantly, RTs were not significant between GP and GP_{NS} ($t(18) = 2.074$, $p = 0.053$, $d = 0.203$). These results suggested that participants prepared the articulation based on the visual cues and the corollary discharge was available since the general preparation stage.

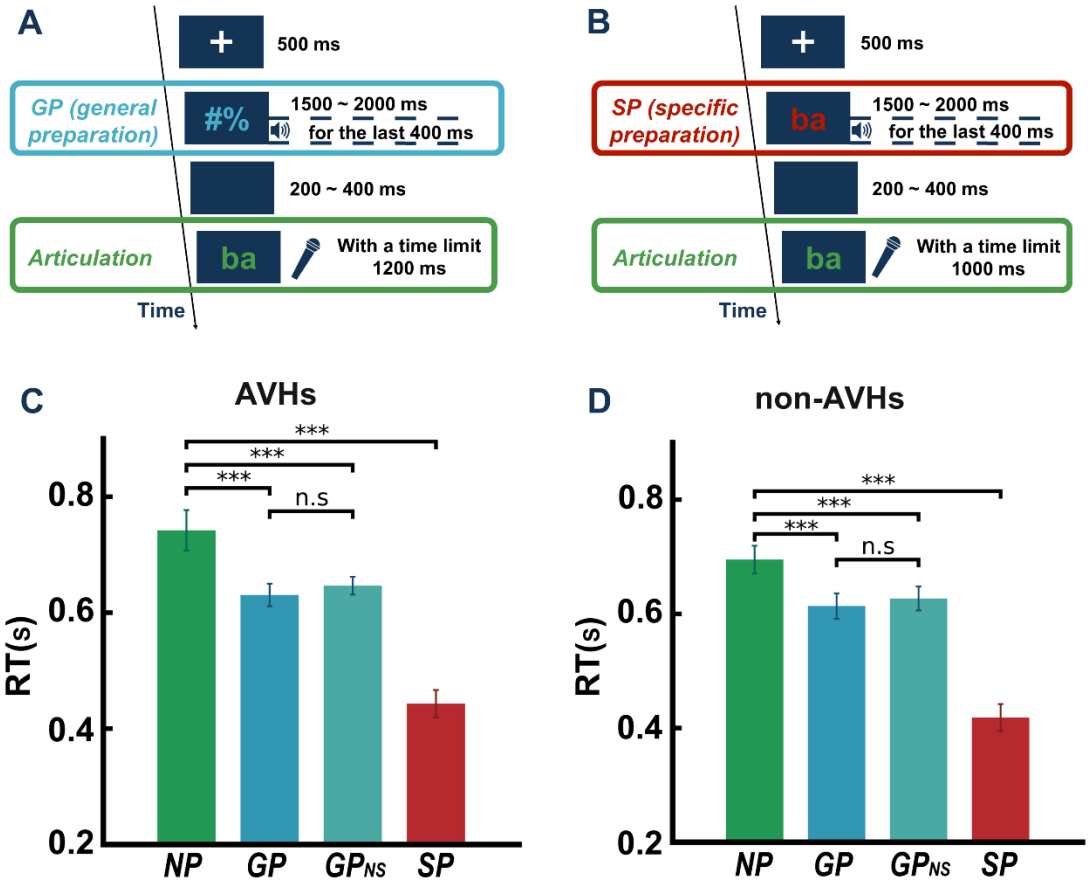

Figure 2. Experimental paradigm and behavioral results in AVHs and non-AVHs patient groups. A) Illustration of a sample trial of general preparation (GP). After a fixation displayed for 500ms, a yellow visual cue of two symbols (#%) appeared in the center of the screen for a range of 1500ms to 2000ms with an increment of 100ms. Participants were asked to prepare to speak in the upcoming articulation task, although they did not know what to say because the symbols did not contain any linguistic information. In half of the trials, an auditory probe, either one of the four auditory syllables (/ba/, /pa/, /ga/, and /ka/) or a 1k Hz pure tone, was played during the last 400ms of the preparatory stage. Another half of the trials did not include any auditory probes (GP_{NS}). After a blank screen with a range of 200ms to 400ms with an increment of 50ms, participants saw a green visual cue that was one of the four syllables (/ba/, /pa/, /ga/, and /ka/) in the center of the screen for a maximum of 1200ms and were asked to produce the syllable as fast and accurately as possible. B) Illustration of a sample trial of specific preparation (SP). The procedure was similar to the GP task with two exceptions: 1) the visual cue during the preparatory stage was a red syllable randomly selected from the four syllables (/ba/, /pa/, /ga/, and /ka/), and 2) an auditory probe was presented in every trial during the preparatory stage. The auditory probes were either the same as or different from the visual cue, yielding two conditions — auditory syllables were congruent (SPcon) or incongruent (SPinc) with the syllable that participants prepared to speak. C) & D) Behavioral results of AVHs and non-AVHs patients. The speed of articulation was measured as reaction time (RT). In both groups, the RTs in GP (with or without auditory probes) and SP were significantly faster than those in NP, suggesting that preparation was carried out in both groups and in all conditions. Error bars indicate ±SEM. $**P < 0.01$, $***P < 0.001$.

In the non-AVHs group (Fig.2D), the statistical results were similar to the ones in the AVHs group. A repeated-measure one-way ANOVA on RTs found a significant main effect of preparation ($F(3,54) = 126.334$, $p < 0.0001$, partial $\eta 2 = 0.894$). Further analysis revealed that the onset of articulation was consistently faster after preparation. Specifically, RTs in SP (mean RT = 418.09 ms; SD = 23.87) were fastest among all four conditions ($F(3,60) = 25.265$, $p < 0.0001$, partial $\eta 2 = 0.558$). RTs in GP (mean RT = 613.56 ms; SD = 22.39) was faster than in NP (mean RT = 695.02 ms; SD = 24.25) ($t(15) = 6.077$, $p < 0.0001$, $d = 0.845$). RTs were also significantly shorter in GP_{NS} (mean RT = 626.93 ms; SD = 21.04) than in NP ($t(15) = 6.576$, $p < 0.001$, $d = 0.726$). These results suggested that participants engaged in speech preparation. Moreover, the RTs were not significantly different between GP and GP_{NS} ($t(15) = 1.511$, $p = 0.152$, $d = 0.149$). These results suggested that participants performed the GP task according to the visual cues and *CD* was available in the general preparation stage. These consistent behavioral results confirmed that both groups of patients can perform the behavioral preparation tasks.

The impaired function of motor signals during general preparation

We first performed within-subject statistical analyses of paired t-tests on the ERPs to the auditory probes in GP to investigate the modulatory effects of CD signals on auditory processes in each patient group. In the AVHs group, the N1 and P2 topographies showed typical auditory response patterns in both GP and B conditions (Fig.3A). However, the magnitude of neural responses in GP was not significantly different from B, neither in the early auditory responses of N1 component ($t(18) = 1.777$, $p = 0.258$, $d = 0.295$) nor in the later auditory responses of P2 component ($t(18) = 0.806$, $p = 0.517$, $d = 0.088$) (Fig.3B). These results contrasted with the ones obtained in the normal population in which GP suppressed auditory responses (Li et al., 2020). These results supported the hypothesis that the function of CD was impaired in AVHs patients.

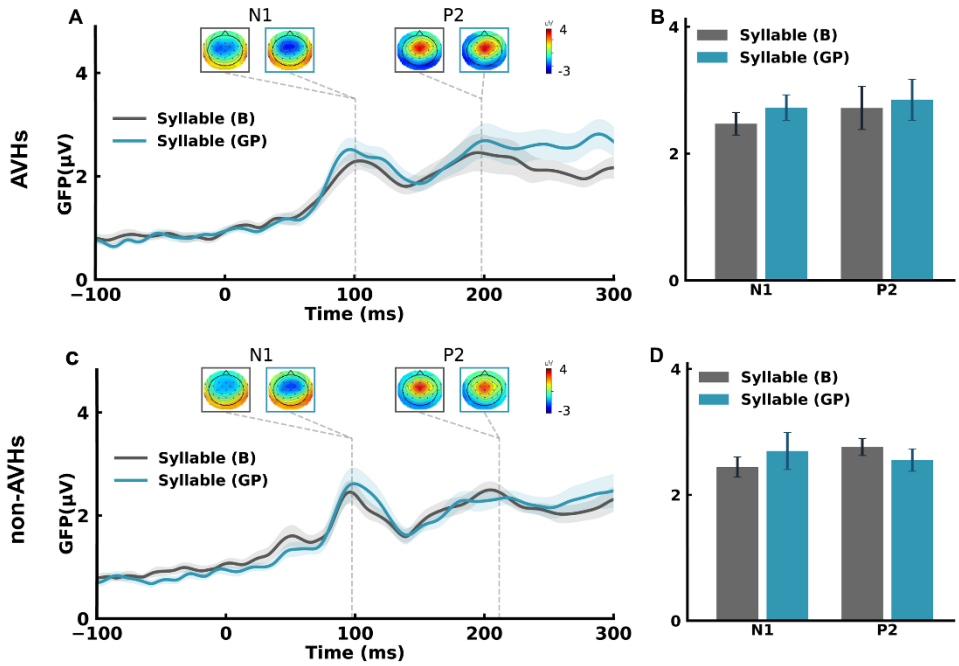

Figure 3. The absence of modulation effects on auditory responses during general preparation (GP) in both AVHs and non-AVHs groups. (A) ERP time course and topographic responses for GP and B conditions in AVHs patients. Peak amplitudes and latencies of the N1 and P2 components were observed in the GFP waveform for each condition. The response topographies at each peak latency are shown in boxes with the same color-code of each condition. (B) Mean GFP amplitudes at N1 and P2 latencies in GP (blue) and B (gray) conditions in AVHs patients. (C) ERP time course and topographic responses in GP and B conditions in non-AVHs patients. (D) Mean GFP amplitudes at the N1 and P2 latencies in GP (blue) and B (gray) conditions in non-AVHs patients. No significant differences between GP and B were observed in either group. Error bars indicate ± SEMs.

In the non-AVHs group, the results were similar to those in the AVHs group. The N1 and P2 topographies showed typical auditory response patterns in both conditions (Fig.3C). The magnitude of neural responses in GP was not significantly different from B, neither in N1 ($t(15) = 1.02$, $p = 0.494$, $d = 0.26$) nor in P2 ($t(15) = -1.166$, $p = 0.332$, $d = -0.318$) (Fig.3D). These results suggest that the effects of CD during the earliest stage of motor intention were also absent in non-AVHs patients.

We further performed two-way mixed ANOVAs to assess the differences in N1 among groups. The group (AVHs, non-AVHs, and normal) was a between-subject factor, and the condition (B and GP) was a within-subject factor. The main effect of group was not significant ($F(2,51) = 1.050$, $p = 0.357$, partial $\eta 2 = 0.040$), neither was the main effect of condition ($F(1,51) = 0.642$, $p = 0.427$, partial $\eta 2 = 0.012$). Whereas the interaction of group and condition was significant ($F(2,51) = 3.416$, $p = 0.041$, partial

$\eta 2$ = 0.118). The significant interaction was further explored by separate ANOVAs for each pairwise group comparison (AVHs vs normal; non-AVHs vs normal and AVHs vs non-AVHs). The group and condition interaction was significant for the comparison of AVHs and normal (F(1,36) = 9.983, p = 0.003, partial $\eta 2$ = 0.217), and non-AVHs and normal (F(1,33) = 4.413, p = 0.043, partial $\eta 2$ = 0.118). However, no significant interaction was observed for the comparison between AVHs and non-AVHs (F(1,33) = 0.000046, p = 0.995, partial $\eta 2$ = 0.000001). These results statistically supported the absence of CD inhibitory effects during the earliest stage of action (i.e. motor intention, manifested by general preparation) in both AVHs and non-AVHs groups.

To explore the modulation functions of CD signals on tones, we conducted paired t-tests on the auditory responses to tones between GP and B for N1 and P2 separately. In the AVHs group, the effects were not significant neither in N1(t(18) = 1.056, p = 0.839, d = 0.124) nor in P2 (t(18) = 0.204, p = 0.84, d = 0.036) (Fig.S1 B). In the non-AVHs group, the effects in N1 (t(15) = 0.735, p = 0.71, d = 0.176) and P2 (t(15) = 0.372, p = 0.715, d = 0.125) were not significant (Fig.S1 D). The N1 and P2 topographies showed typical auditory response patterns in both groups and conditions (Fig.S1 A & Fig.S1 C). The absence of modulation effects on tones in AVHs and non-AVHs patients contrasted with the increased error responses of N1 in normal controls, further supporting the hypothesis of impaired CD during motor intention in schizophrenia patients.

The functions of motor signals during specific preparation dissociated between AVHs and non-AVHs

We next investigated the function of EC by paired t-tests on the ERPs to the auditory probes in SP. In the AVHs group, the N1 and P2 topographies showed typical auditory response patterns among SPinc, SPcon and B conditions (Fig. 4A). The magnitude of N1 was larger than that in B when the auditory syllables were incongruent with the contents of specific preparation (SPinc) (t(18) = 2.242, p = 0.038, d = 0.488). The effect was not significant in the later auditory responses of P2 (t(18) = -1.32, p = 0.407, d = -0.163). However, when the auditory syllables were congruent with the specific

preparation (SPcon), the effect was not significant in N1 ($t(18) = 0.967$, $p = 0.346$, $d = 0.199$), nor in P2 ($t(18) = -0.14$, $p = 0.89$, $d = -0.018$) (Fig. 4B). These results were opposite to the results in normal controls in which motor signals during SP enhanced the perceptual responses to the congruent auditory syllable probes (Li et al., 2020). The observed modulation effects in AVHs suggested that EC was available during specific preparation, but the opposite modulation patterns (enhancement in SPinc in AVHs compared with enhancement in SPcon in normal) indicated a 'noisy' EC that yield imprecise modulation on auditory responses during specific preparation in AVHs.

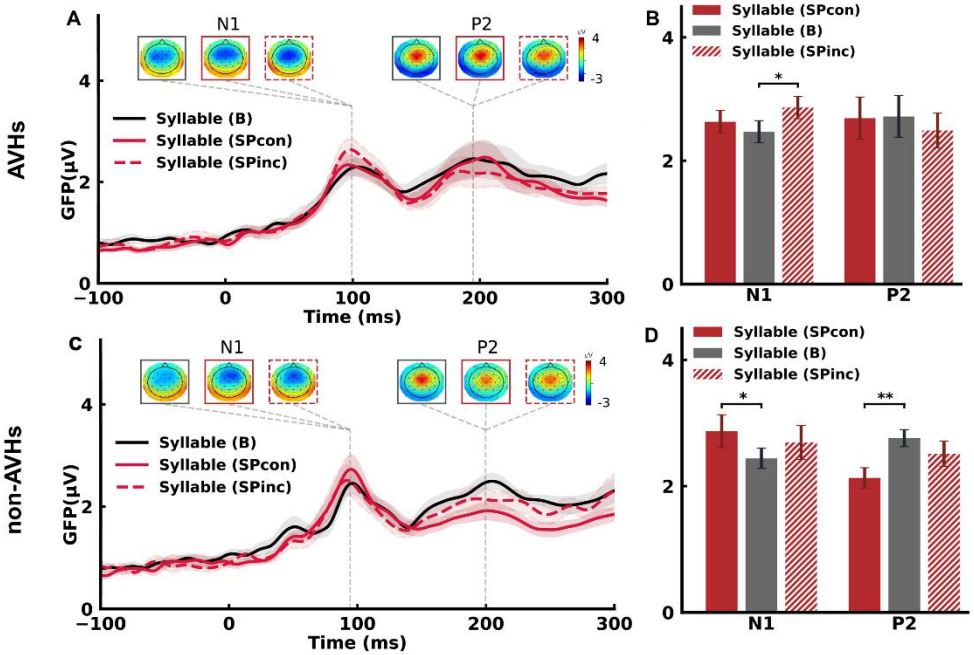

Figure 4. The opposite modulation effects on auditory responses during specific preparation (SP) between AVHs and non-AVHs groups. (A) ERP time course and topographic responses for SP and B conditions in AVHs patients. Typical N1 and P2 auditory response components were observed in GFP waveforms of each condition. The response topographies at each peak latency are shown in boxes with the same color-code of each condition. (B) Mean GFP amplitude at N1 and P2 latencies for SP (red) and B (gray) conditions in AVHs patients. Responses in SPinc were significantly larger than those in B in N1 component. (C) ERP time course and topographic responses for SP and B conditions in non-AVHs patients. (D) Mean GFP amplitudes at N1 and P2 latencies for SP and B conditions. Responses in SPcon were significantly larger than those in B in N1 components, contrasting with the results in AVHs in B). Error bars indicate ±SEMs. ∗ for $p < 0.05$, FDR-corrected for multiple comparisons.

In the non-AVHs group, the N1 and P2 topographies showed typical auditory response patterns in all conditions (Fig.4C). When the auditory syllables were congruent with the specific preparation (SPcon), the response magnitude of the N1 component was larger than that in B ($t(15) = 2.90$, $p = 0.011$, $d = 0.482$), whereas the

response magnitude of P2 component was reduced relative to B ($t(15) = -3.67$, $p = 0.009$, $d = -1.024$). In SPinc, the response magnitude of N1 ($t(15) = 1.24$, $p = 0.234$, $d = 0.273$) and P2 ($t(15) = -1.129$, $p = 0.332$, $d = 0.351$) was not significantly different from B (Fig.4D). The results of SP in non-AVHs were consistent with the results from the normal control group (Li et al., 2020), suggesting an intact EC function in non-AVHs.

For responses to tones, paired t-tests were conducted between SP and B for N1 and P2 separately. The effects were not significant either in N1 ($t(18) = 0.558$, $p = 0.839$, $d = 0.078$) or in P2 ($t(18) = -0.771$, $p = 0.676$, $d = 0.116$) in the AVHs group (Fig.S2 B). In non-AVHs group, the effects in N1 ($t(15) = -0.039$, $p = 0.969$, $d = -0.007$) and P2 ($t(15) = -1.09$, $p = 0.59$, $d = -0.234$) were not significant (Fig.S2 D). The N1 and P2 topographies showed typical auditory response patterns in both conditions and groups (Fig.S2 A & Fig.S2 C). The lack of modulation on tones during specific preparation in both patient groups was consistent with the results in normal controls, indicating that EC contained the task-related information in both AVHs and non-AVHs.

The correlation of clinical symptoms and neural modulation effects

Pearson correlation analyses were performed to investigate the relationship between the neural responses and the clinical symptoms. According to the hypothesis, the less severe the AVH symptom in the AVHs group, the more similar their neural modulation effects would be to the normal controls. A difference score was calculated by subtracting the N1 response amplitude in the B condition from the SPcon condition in the AVHs group, representing the magnitude of the N1 enhancement effect in the SPcon condition. We found that the N1 enhancement magnitude significantly and negatively correlated with the positive symptoms total scores of the PANSS (Fig.S3, $r = -0.5726$, $p = 0.0104$). That is, the more severe the positive symptoms, the less enhancement in the N1 responses. The results suggest that no significant enhancement of N1 response magnitude in SPcon condition in AVHs groups was related to the degree of positive symptom severity.

Modeling results of dissociative impairment of CD and EC in AVHs and non-AVHs

To collaboratively simulate the dissociations of impairment between the effects of CD

and EC in AVHs and non-AVHs groups, we quantified our hypotheses in a two-layer neural network model. The upper motor layer had two functional pathways that linked to the lower auditory layer: an indirect pathway via an interneuron to inhibit all nodes in the auditory layer, and a direct pathway for modulating the gain of corresponding auditory nodes (Fig. 5A). These indirect inhibitory and direct enhancement pathways manifest the functions of CD and EC, respectively. This bifurcation of motor signals successfully explained the distinct modulation directions in GP (Fig. 5B, left) and SP (Fig. 5C, left) in normal participants (Li et al., 2020).

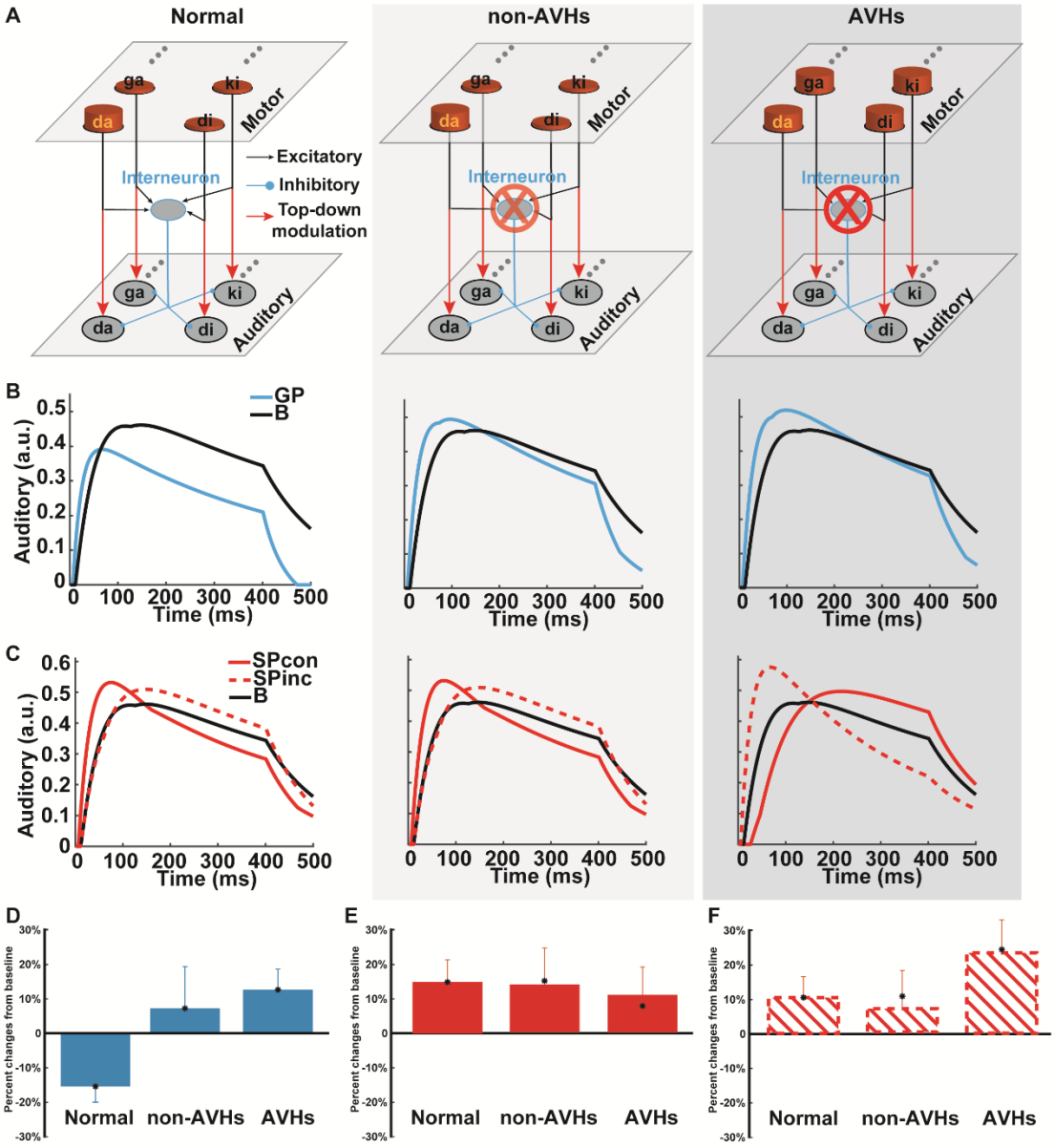

Figure 5. Model simulation results of distinct impairments of CD and EC in clinical populations. A) model architecture and manipulations that quantify the hypothesized neural impairments mediating AVH.

Speech units in the motor layer (upper) link to the units in the auditory layer (lower) via two pathways. An interneuron receives signals from each unit in the motor layer and inhibits all units in the auditory layer (blue, simulating the inhibitory function of CD). Moreover, signals from each motor unit bifurcate and sensitize its corresponding auditory unit (red arrows, simulating the enhancement function of EC). For non-AVHs (middle plot), the weaker CD function is modeled as the reduced inhibition strength of the interneuron (indicated by a lightened red cross). Whereas the modulation strength of the EC enhancement function is preserved in non-AVHs, indicated by the only red column on the syllable 'da' (taking a trial that 'da' is prepared for an example). The selective modulation is available on the prepared syllable, identical to the function in normal controls (left). For AVHs (right plot), the inhibitory function of CD could be even more impaired (a red cross over the interneuron) and the enhancement function of EC is imprecise as its modulation not only on the target unit of 'da' but also over all neighboring units (elevated columns in all units). The simulation results of time course responses in B) GP and in C) SP conditions for normal (left), non-AVHs (middle), and AVHs (right) groups. The simulation results of component response magnitude compared with the empirical data in D) GP, E) SPcon, and F) SPinc. The bars represent the observed modulation effects of speech preparation on N1 auditory responses. The stars on the bars represent the simulation results in a given condition and group. Both empirical and simulation results are normalized by baseline conditions. The model simulations capture the absence of inhibition during GP and opposite modulation patterns in SP among different groups.

For the non-AVHs group, we set the inhibition strength of the interneuron as a free parameter, while other parameters remained the same as the previous study of fitting the results of the normal population. If only CD at the movement intention stage was impaired but EC at the specific preparation stage was intact in the non-AVHs group, the inhibition strength of the interneuron would be significantly smaller than that of the normal population, whereas the same gain modulation in the direct pathway that used in fitting normal population would be able to fit the results of non-AVHs (Fig. 5A, middle). More importantly, we hypothesized a 'broken' CD and a 'noisy' EC for the AVHs group. The 'broken' CD hypothesis led to a similar prediction that the inhibitory strength of the interneuron would be smaller than the normal population, and even smaller than non-AVHs (Fig. 5A, right). To test the hypothesis of 'noisy' EC, we manipulated the ratio of the gain modulation strength between the prepared and unprepared syllables. The 'noisy' EC hypothesis derived a prediction that the best-fitted parameter for the SP results in AVHs would yield relatively more gain over the neighboring auditory nodes than that of the prepared auditory target (Fig. 5A, right).

For the non-AVHs group, the simulation results revealed that, during GP, the inhibitory effect was dampened because of deficits in the interneuron. The best-fitted

parameter of inhibitory strength was 0.1711, compared with stronger inhibition of 0.4372 in normal population. The weaker inhibitory strength made the inhibition of GP disappear (Fig. 5B, middle). After temporal averaging of the peak component in the waveform responses, the simulation result of suppression in GP relative to B was consistent with the empirical observations in the *GP* condition for non-AVHs group (Fig. 5D). The Bayes factor of comparison between the simulation results and empirical results in *GP* condition favored the null (scaled JZS Bayes factor = 3.92), suggesting the model captured the absence of suppressed auditory responses in the *GP* for non-AVHs group.

For AVHs group, a similar inhibitory deficit was built in the interneuron (Fig. 5A. right). The simulation results suggest that during GP, the inhibitory strength was decreased to 0.1259, much smaller than that of the normal population (0.4372), and smaller than that of the non-AVHs group (0.1711). The much weaker inhibitory strength yielded the absence of inhibition in GP (Fig. 5B, right). The difference between temporal averages of the simulated peak components of GP and B in the AVHs group was not different from the empirical results, supported by the Bayes factor (scaled JZS Bayes factor = 4.21), suggesting the model also captured the absence of suppression in the GP for AVH group.

The modulation effects diverged between AVHs and non-AVHs groups in SP -- the non-AVHs group had similar effects as the normal population, whereas AVHs had an opposite modulation pattern (empirical results in Fig. 4). This was hypothesized as the EC function differences – 'noisy' in AVHs group, whereas the EC function remained as precise in the non-AVHs group as normal group. For the simulation of SP results in non-AVHs, the gain modulation function remained the same as that in the normal population, indicated by a similar pattern of columns on motor units between non-AVHs (Fig. 5A, middle) and normal (Fig. 5A, left). The simulation results showed the enhancement effects in SP compared to B in the non-AVHs group (Fig. 5C, middle). The temporal averages around the peak of simulated waveform responses were not statistically different from the empirical data, supported by the Bayes factor (scaled JZS Bayes factor = 3.90 for SPcon, and 3.74 for SPinc), suggesting the intact modulation of

EC in non-AVHs. Interestingly, the simulation results of SP in non-AVHs must use the normal inhibitory strength of the interneuron rather than the simulated decreased inhibitory strength obtained in GP of non-AVHs. This indicated that the inhibitory function may vary across the preparation stages that potentially originated from different hierarchies of the neural pathway for actions (Li et al., 2020) and the deficits of CD in the non-AVHs group may be only in the early intentional stage but not in later preparatory stages.

For AVHs group, in addition to the inhibitory deficits in the interneuron, the gain modulation function was hypothesized to be 'noisy'. The 'noisy' gain modulation was modeled as a parameter of the ratio between the modulation gain on prepared and unprepared auditory nodes. That is, the gain modulation in AVHs may not as precise as that in normal or non-AVHs groups (indicated in Fig. 5A right plot, the columns on the motor units for the unprepared syllables were also increased in AVHs compared to the little magnitude for the unprepared syllables in the normal and non-AVHs group in Fig. 5A left and middle). The simulation was performed simultaneously for the SPcon and SPinc results in AVHs. The best-fitted parameter of the ratio was 2.286, yielding the modulation gain between prepared and unprepared nodes of 0.437 : 2.286, compared with that in normal and non-AVHs groups of 1.919 : 1. This more 'spread-out' modulation gain to the neighboring nodes of the prepared target resulted in smaller responses to SPcon but stronger responses to SPinc as compared to B (Fig. 5C). The temporal averages around the peak of simulated waveform responses were statistically not different from the empirical data, as supported by the Bayes factor (scaled JZS Bayes factor = 3.91 for SPcon, and 4.19 for SPinc), suggesting the imprecise modulation by a 'noisy' EC in AVHs.

Discussion

We investigated the modulation functions of motor signals on auditory processing at distinct stages of speech preparation in schizophrenia patients. Our behavioral, electrophysiological, and modeling results collaboratively demonstrated distinct impairments in the motor-to-sensory transformation between subgroups of clinical

populations. The symptoms of AVHs are mediated by the impaired source monitoring function of *CD* that results in the absence of inhibition of auditory responses in the general speech preparation, as well as the imprecise activation function of *EC* that results in the varied enhancement and sensitization of auditory cortex during specific speech preparation. These results suggest that 'broken' *CD* plus 'noisy' *EC* causes erroneous monitoring of the imprecise generation of internal auditory representation and hence yields AVHs.

The absence of suppression effects during general preparation in schizophrenia patients is the negative evidence supporting that *CD* signals can be generated during the movement intention stage even without any preparatory contents. Compared with the function of ubiquitously suppressing neural responses to speech in normal controls (Li et al., 2020; Zheng et al., 2022), the *CD* did not function in both groups of schizophrenia patients with and without AVHs at the earliest stage of motor intention (Fig. 3). These results were consistent with immense literature about the weaker or absent action-induced sensory suppression in schizophrenia (Ford et al., 2001; Ford et al., 2001; Ford et al., 2014; Ford et al., 2007; Whitford, 2019), as well as findings that schizophrenia patients elicit a smaller readiness potential before movement (Deecke et al., 1969) than normal controls (Ford et al., 2014; Pinheiro et al., 2020). Our results delineate the temporal dynamics of the impaired function of *CD* that can occur in the earliest stage of motor intention, complementing with and extending from most findings during the action execution.

Different causes may mediate the absence of inhibition effects during general preparation in distinct clinical subgroups. Lack of desire to act may be a common cause in both patient groups that decreases the inhibitory strength of *CD* on auditory processes (Fig. 3) because no significant differences were found in the severity of their negative symptoms. The ubiquitous pathophysiological negative symptoms in schizophrenia may generate weaker *CD* in the motor intention stage in both patient groups. That is, the lack of inhibitory effects during general action preparation may be a biomarker for less intention and negative symptoms. However, impaired *CD* functions in motor-to-sensory transformation may be a cause that unique in AVHs in addition to the deficits

in the generation of *CD* due to negative symptoms. Although no difference in the negative empirical results of modulation by general preparation (Fig. 3), computational model results (Fig. 5) reveal 1) a greater degree of impairment in the inhibition strength of the interneuron in AVHs compared with non-AVHs; and 2) the impaired function of *CD* continued throughout the specific speech preparation stage in AVHs, whereas the empirical results of modulation in SP in non-AVHs require the intact *CD* inhibitory strength to fit. This evidence regarding the degree and temporal extent of the inhibitory function deficits suggests the impairment of *CD* in the motor-to-sensory transformation in AVHs. Our results of more severe impairment of *CD* in general preparation and throughout specific preparation in AVHs but not in non-AVHs are consistent with the findings that transcranial magnetic stimulation affected the sense of agency only when stimulation time locked in action planning, rather than in the physical consequences of the actions appeared (Zapparoli et al., 2020). The empirical and modeling results consistently support that the impairment of *CD* function associates with an abnormal sense of agency in AVHs patients (Feinberg & Guazzelli, 1999).

The *EC* function and its impairment also show dissociation between AVHs and non-AVHs. In non-AVHs patients, the *EC* function is the same as normal -- the motor signals in specific preparation enhanced the neural responses only to the prepared syllable (Fig. 4). However, in AVHs patients, the motor signals in specific preparation enhanced the neural responses to the unprepared syllable (Fig. 4). The modulation effects of enhancement are negatively correlated with the severity of AVH symptoms (Fig. S3). Modeling results further quantified that the different modulation patterns between AVHs and non-AVHs were caused by the imprecise modulation from the motor to sensory units that provide incorrect gains over the non-target of the specific preparation (Fig. 5). These results suggest that *EC* can be generated in the motor-to-sensory transformation pathway during specific preparation in AVHs patients. However, the EC is 'noisy' either in the generation process in the motor system or it is imprecisely mapped onto the auditory system. As a result, the 'noisy' EC modulates and enhances the sensitivity of neural responses to unprepared auditory units. The empirical ERP modulation effects, correlation results with AVH symptoms, and model simulation

results collaboratively support the hypothesis of 'noisy' EC in AVHs. This imprecise EC in AVHs may relate to the 'non-sense' and various forms of auditory hallucinations.

In previous studies, the action-induced suppression (Blakemore et al., 1998; Crapse & Sommer, 2008; Houde & Nagarajan, 2011; Miall & Wolpert, 1996) and enhancement (Eliades & Wang, 2005, 2008; Enikolopov et al., 2018; Flinker et al., 2010; Singla et al., 2017) have been observed in the normal population and animal models. By considering the different characteristics of motor signals across temporal dynamics of motor processes, it has been proposed that the copy of motor signals at different action stages may distinguish into *CD* and *EC* that mediate distinct functions for regulating actions (Li et al., 2020). In this study, we found distinct impairments in AVHs between general and specific preparation stages (Figs. 3&4). The observed distinct impairments in different speech preparation stages offer evidence from a clinical perspective that is consistent with empirical neuroscience results in the normal population and support the updated theoretical framework of internal forward models (Li et al., 2020).

The observed double dissociations of *CD* and *EC* functions between AVHs and non-AVHs reveal the impairments in the motor-to-sensory transformation that mediate the positive symptoms of auditory hallucinations. The positive nature of auditory hallucinations requires the active construction of neural representations that mediate perceptual-like experience. Our observation of 'noisy' *EC* that imprecisely sensitizes auditory cortices provides a foundation for inducing subjective experience without external stimulation. Together with the impairment in *CD* that leads to less suppression and hence deficits in labeling the sources that induce neural responses, hallucinations about experiencing perceptual events would occur. That is, the combination of impairments on distinct functions between motor and sensory systems mediate the positive symptoms of auditory hallucination, which is consistent with the hypothesis of motor-to-sensory transformation as an origin of hallucinations (Ford et al., 2001; Shergill et al., 2000; Stephan et al., 2009; Toh et al., 2022; Yang et al., 2019) – the impaired monitoring function misattributes the sources of internally motor-induced (Li et al., 2020) or other top-down induced neural representations (Bansal et al., 2022; Yang et al., 2021). The conceptual, anatomical, and functional distinct motor signals of *CD*

and *EC*, instead of sole inhibitory function in the motor-to-sensory transformation, collaboratively contribute to the positive symptoms of auditory hallucinations.

Most previous studies have explored the differences in impaired motor-to-sensory transformation signals between schizophrenia patients and normal controls (Blanchard & Neale, 1994; Feinberg & Guazzelli, 1999; Ford et al., 2008; Heinks-Maldonado et al., 2007). In this study, we explored how the motor signals regulated perceptual neural responses between subgroups of schizophrenia patients with different symptoms. This hypothesis-driven symptom-based approach yields novel insights into the potential neural mechanisms that mediate different aspects of deficits in schizophrenia. By distinguishing the uniqueness of psychotic symptoms in the same categorized mental disorder, the distinct impairments in the motor-to-sensory transformation have been revealed, which delineates the potential deficits mediating positive symptoms in mental disorders. Moreover, considering the overlapping symptoms between subgroups of patients can provide insights into possible causes, for example, the negative symptoms in both AVHs and non-AVHs may reveal the deficits in the motor system in relation to anhedonia and amotivational syndrome. The approach of comparing unique and common symptoms may expand over different types of mental disorders, such as auditory hallucinations in schizophrenia and bipolar disorder to investigate the potential common causes from a cognitive neuroscience perspective. Adding the neural bases of symptoms across different types of mental disorders complements the symptoms-based categorization and may provide a 2-D matrix for more precise diagnosis of mental disorder (Gold et al., 2020).

Our study highlights cognitive computation as a crucial interface to bridge neural circuits to mind and behavior, especially in understanding mental disorders. Recently, computational psychiatry has emerged as a novel quantitative cognitive account for probing the mechanisms that mediate mental disorders (Friston et al., 2014; Huys et al., 2016; Montague et al., 2012; Wang & Krystal, 2014). In this study, we utilized a computational modeling approach and identified the subtle differences in the negative results during general preparation between AVHs and non-AVHs. Moreover, the computational modeling enables us to differentiate the distinct impairments of *CD* and

EC throughout the evaluation of actions – parametric simulation using hypothesized intact and impaired values overcome the temporal overlaps of *CD* and *EC* in the specific preparation that would be hard, if not impossible to investigate using behavioral or non-invasive cognitive neuroscience approaches on human participants with mental disorders. The consistent EEG and modeling results mechanistically reveal the predictive functions of motor and sensory networks that may mediate the symptoms of psychosis (Fletcher & Frith, 2009; Stephan et al., 2009). Our endeavors of combining behavior, electrophysiology, and modeling manifest Marr's computational approach (Marr, 2010) and provide a possible link between mental and behavioral status with neural circuits (Carandini, 2012; Marcus et al., 2014). The computational approach puts the cognition back to the investigation of mental disorders (Taschereau-Dumouchel et al., 2022), and yields testable hypotheses at the cognitive, system and even cellular and molecular levels to collaboratively understand mental disorders.

By probing the impairments in the interactive neural processes between motor and sensory systems, we observed the functional distinctions between *CD* and *EC*, and their impairments in relation to auditory hallucinations in schizophrenia. The pathophysiology of schizophrenia involves a ubiquitously distributed motor-sensory circuitry in which the 'broken' *CD* dysfunctionally misattributes the sources of the neural activity induced by 'noisy' EC – the failure of dampening the internally induced sensory neural activity leads to hallucinatory experiences. Distinct impairments in functional granularity of motor-to-sensory transformation mediate positivity symptoms of agency deficits in mental disorders.

Materials and Methods

Participants

Nineteen patients (nine males), who matched the DSM-V diagnosis of schizophrenia and were experiencing AVHs without hallucinations in other modalities (AVHs group), were recruited from Shanghai Mental Health Center. Another group of sixteen patients (ten males), who met a DSM-V diagnosis of schizophrenia and had never experienced

AVHs (non-AVHs group), were recruited from the same hospital. Two experienced psychiatrists independently confirmed the diagnoses based on the Structured Clinical Interview for DSM-V. All patients were right-handed with an age range of 18-45 years old. They received antipsychotic medications and were clinically stable during the experiment. The study was approved by the Institutional Review Board at New York University Shanghai and the Institutional Ethics Committee at Shanghai Mental Health Center. Patients provided written informed consent before they participated in the study. This study was performed in accordance with the guidelines laid out in the Helsinki Declaration of 1975, as revised in 2008.

Previous results (Li et al., 2020) of nineteen normal participants who completed the same GP task as the patients and sixteen normal participants who completed the same SP task served as a normal control group.

Clinical measures

All participants' demographic data were collected and reported. The duration of illness of each patient was recorded. Symptom rating interviews were done on the day of EEG recordings. Trained psychiatrists assessed the psychotic symptoms using the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987). The PANSS measures the presence and severity of positive, negative, and general symptoms on a seven-point scale. Hallucination severity was rated from the P3 subscore of the PANSS, with the higher rating indicating an increased hallucination severity. Non-AVHs patients had a rating of 1 in the P3 subscore, indicating the hallucination symptom was absent. Furthermore, the severity level of AVHs was assessed using the seven-item Auditory Hallucinations Rating Scale (AHRS) (Hoffman et al., 2003).

Materials

Four auditory syllables (/ba/, /pa/, /ga/, /ka/) with a duration of 400ms were synthesized via the Neospeech engine (www.neospeech.com) at a 44.1kHz sampling rate in a male voice. Moreover, a 1k Hz pure tone with the same duration of 400 ms was included in the experiment. All auditory stimuli were delivered binaurally at 70dB SPL through plastic air tubes connected to foam earplugs (ER-3C Insert Earphones; Etymotic

Research). A Shure Beta 58A microphone was used to detect and record participants' vocalization.

Procedures

We first provide an overview of the procedures. Two tasks were used to elicit different speech preparation stages before articulation. This design provided the temporal segregation of motor signals and induced corollary discharge (CD) and efference copy (EC) in different stages of motor preparation. During each preparation stage, auditory probes were introduced to explore how distinct preparation stages (and hence different motor signals of CD and EC) modulate perceptual responses to auditory stimuli. Below we describe the details of the procedures.

Fig. 2A illustrates the trials of the general preparation (GP) task. The trial started with a cross fixation displayed for 500ms. A yellow visual cue of two symbols (#%) then appeared in the center of the screen for a range of 1500ms to 2000ms with an increment of 100ms. Participants were asked to prepare to speak in the upcoming articulation task, although they did not know what to say because the symbols did not contain any linguistic information. In half of the trials, an auditory probe, either one of the four auditory syllables (/ba/, /pa/, /ga/, and /ka/) or a 1k Hz pure tone, was played during the last 400ms of the preparatory stage. Another half of the trials did not include auditory probes (GP_{NS}). The mixed-trial design aimed to enforce participants' preparation for the upcoming articulation task based on visual cues instead of auditory probes. After a blank screen of a range of 200ms to 400ms with an increment of 50ms, participants saw a green visual cue that was one of the four syllables (/ba/, /pa/, /ga/, and /ka/) in the center of the screen for a maximum of 1200ms, and were asked to produce the syllable as fast and accurately as possible. The onset time of vocal response was recorded to quantify the reaction time (RT).

Fig. 2B illustrates the trials of the specific preparation (SP) task. The procedure was similar to the GP task except for two differences. One was that the visual cue during the preparatory stage was a red syllable randomly selected from the four syllables (/ba/, /pa/, /ga/, and /ka/). Participants prepared to speak the syllables because the upcoming

articulation task was the same syllable with a speeded requirement in a time-limited setting. The other difference was that an auditory probe was presented in every trial during the preparatory stage. The auditory probes were either the same as or different from the visual cue, yielding two conditions — auditory syllables were congruent (SPcon) or incongruent (SPinc) with the visual cue (and hence the syllable that participants prepared to speak).

Furthermore, two additional types of trials were included to reduce expectations and to evaluate the effects of preparation. In one type of trials, the green visual cue of articulation immediately appeared after the fixation, and participants were asked to articulate the syllable without preparation (NP trial). The RTs in NP trials were used as a behavioral baseline of syllable production speed and were compared with the RTs in preparation trials to quantify the effects of preparation behaviorally. In another type of trials, a white visual cue of symbols (**) appeared after the fixation with auditory probes played during the last 400ms of the visual cue presentation duration. No articulation stage followed. Participants were only required to listen to the auditory probes passively (baseline, B trial). The B trials contained similar visual cues and auditory probes as those in the GP and SP trials but without preparation, yielding baseline auditory responses for quantifying the neural modulation effects of preparation. The neural responses to the auditory probes in the B trials were compared to those in GP and SP trials to quantify the modulation effects of different motor signals. The NP and B trials were equally divided into GP and SP blocks to reduce the duration of the experiment.

Therefore, four types of trials (NP, GP, GP_{NS}, B) were randomly presented within six GP blocks. Each block included 48 trials, yielding a total of 288 trials (96 trials for GP_{NS} and GP; 48 trials for NP and B). Half of the GP trials and half of the B trials contained auditory probes of syllables and another half contained auditory probes of pure tone. The time limit for articulation was set to 1200ms for GP and GP_{NS} trials, respectively. The time limit setup was aimed to eliminate expectations and enforce preparation.

Three types of trials (NP, SP, B) were randomly presented within five SP blocks. Each block included 48 trials, yielding 240 trials (144 trials for SP; 48 trials for NP and B). A third of the SP trials were SPcon condition, another third was SPinc condition, and the last third of trials had auditory probes of pure tone. The 48 B trials also contained half trials of auditory syllables and half trials of pure tone. Together with the B trials in GP blocks, a total of 48 trials with auditory syllables and 48 trials with pure tone for the B condition. The time limit for articulation was set to 1500ms for NP, and 1000 ms for SP, respectively.

Demographic and clinical data

Statistical analyses were performed using IBM SPSS (Statistics version 17.0) and GraphPad.Prism 5.02. The normality of data was tested using the Kolmogorov-Smirnov tests. Demographic and continuous variables were subject to one-way ANOVA and Fisher's LSD (least significant difference) post hoc multiple comparison tests ($\alpha = .05$), whereas the categorical values were subject to the chi-squares test. Data were presented as mean and standard deviation. Effects at $p < 0.05$ were considered significant.

Behavioral data analysis

The RTs of the articulation were calculated as the time lag between the onset of the green visual cue and the participants' vocalization. In AVHs and non-AVHs groups, the averaged RTs were obtained in each of the four trial types (NP, GP, GP_{NS}, SP). The RTs were subject to a repeated-measures one-way ANOVA and a post-hoc Tukey Student t-test for pairwise comparisons.

EEG data acquisition and preprocessing

Neural responses were recorded using a 32-channel active electrode system (Brain Vision actiCHamp; Brain Products) with a 1000 Hz sampling rate in an electromagnetically shielded and sound-proof room. Electrodes were placed on an EasyCap, on which electrode holders were arranged according to the 10-20 international electrode system. The impedance of each electrode was kept below 10kΩ. The data were referenced online to the electrode of Cz and re-referenced offline to the

grand average of all electrodes. Two additional EOG electrodes (horizontal: HEOG; vertical: VEOG) were attached for monitoring ocular activity. The EEG data were acquired with Brain Vision PyCoder software (http://www.brainvision.com/pycorder.html) and filtered online between DC and 200Hz with a notch filter at 50Hz.

EEG data processing and analysis were conducted with customized Python codes, MNE-python (https://github.com/mne-tools/mne-python) (Gramfort et al., 2014), Autoreject (https://github.com/repos/autoreject) (Jas et al., 2017), EasyEEG (https://github.com/ray306/EasyEEG) (Yang et al., 2018), and the Topography-based Temporal-analysis Toolbox (TTT, https://github.com/TTT-EEG/TTT-EEG) (Wang et al., 2019). For each participant's dataset, bad channels were replaced with the average of their neighboring channels. The dataset was band-pass filtered with cut-off frequencies set to 0.1 and 30Hz. The filtered dataset was then segmented into epochs ranging from -200ms to 800ms, relative to the onset of the auditory probe, and baseline corrected using the 200ms pre-stimulus period. 240 epochs in the GP task and 192 epochs in the SP task were extracted for each participant. Epochs with peak-to-peak amplitude exceeding the threshold determined by the Autoreject toolbox were automatically excluded. Epochs with artifacts related to eye blinks and head movement were manually rejected. To ensure data quality, we excluded epochs that were contaminated by any residual noise before analysis. On average, 259 epochs were included in the AVHs group and 262 epochs in the non-AVHs group. In the AVHs group, for trials with auditory syllable probes, on average, 32, 35, 40 and 39 trials were included in B, GP, SPcon, and SPinc, respectively. For trials with auditory probe of tones, on average, 36, 36 and 41 trials were included in B, GP, and SP, respectively. In the non-AVHs group, for trials with auditory syllable probes, on average, 36, 39, 39 and 36 trials were included in B, GP, SPcon, and SPinc, respectively. For trials with auditory probes of tones, on average, 32, 40 and 40 trials were included in B, GP, and SP, respectively. The ratio of trial rejection in the AVHs group and the non-AVHs group was 22.74% and 22.05%, respectively.

Event-related potentials (ERPs) to the auditory probes were obtained by averaging

trials in each condition (syllables in GP, SPcon, SPinc, and B; tones in GP, SP, and B) for each participant in the AVHs and non-AVHs groups. The global field power (GFP) — the normalized geometric mean across 32 electrodes — was calculated from each ERP. The GFP represents the overall response power changes over time, which is an optimal measure in a novel study to balance the requirements of exploration and to overcome problems of false positives by avoiding subjective channel selections, multiple comparisons, and individual differences (Tian & Huber, 2008; Tian et al., 2011). Individual peak amplitudes and peak latencies of the N1 and P2 components in the GFP waveforms were automatically identified using the TTT toolbox in predetermined time windows of 50-150 ms and 150-250 ms, respectively (Wang et al., 2019). We visually verified whether identified peaks were correct in each participant.

EEG data analysis

The identified ERP component responses were used in the following statistical tests. First, within-group analyses were performed, separately in the AVHs and non-AVHs groups, to test the hypotheses about the impairment of motor signals and their modulation effects in each group. Paired t-tests were carried out on the GFP responses to syllables in the hypothesis-driven paired comparisons (GP and B, SPcon and B, and SPinc and B). Repeated measures one-way ANOVAs were conducted on GFP responses to tones (GP, SP, and B). Both statistical tests were performed separately for the N1 and P2 components. Furthermore, two-way mixed ANOVAs were conducted to test the between-group differences, separately for the GP, SPcon, and SPinc conditions. The factor of group (AVHs, non-AVHs, and normal) was set as a between-subject factor. The condition (GP and B, SPcon and B, and SPinc and B) was set as the within-subject factor. For ANOVAs, effect sizes were indexed by partial $\eta 2$. For paired t-tests, effect sizes were indexed by Cohen's d. Significant effects were determined by $p < 0.05$ and partial $\eta 2 > 0.14$ (Richardson 2011).

Modeling

We adapted the model in Li et al. (2020) to quantitatively test our hypotheses regarding the deficits of CD and EC in AVHs and non-AVHs groups. Specifically, two free

parameters were introduced in the model to adjust the values that represent the CD and EC functions. The first parameter was *the inhibitory strength* of an intern-neuron to simulate the hypothesis of impaired CD during GP. The second parameter was *the ratio of modulation gain* between prepared and unprepared auditory tokens to simulate the hypothesis of 'noisy' EC during SP in AVHs. Next, we described the models and the incorporation of two parameters to test our hypotheses.

To quantify the proposed double dissociation between the impairment of CD and EC in AVHs and non-AVHs groups, we built a two-layer neural network model to simulate the dynamics and modulation effects of motor signals on sensory processing (Fig. 5A). The model structures are similar to the previous model of simulating CD and EC in the normal participant (Li et al., 2020) (Fig. 5A, left plot). The upper layer represents motor processing, and the lower layer denotes auditory processing. Each layer includes multiple neurons that represent different syllables. Each neuron in the auditory layer is a rate-coded unit with synaptic depression. The updating of membrane potential is governed by Eq. (1).

$$\frac{dv_i(t)}{dt} = \tau\left\{g'_i(1-v_i)\sum_j w_{ij}e_j - v_i[L + I(\sum_k o_k + n * m')]\right\} \qquad \text{(Eq. 1)}$$

The membrane potential of an auditory neuron, v_i, is updated according to the sum of three sources. The first source is an excitatory input from acoustic signals, e_j, via bottom-up connections with connection strength, w_{ij}. This bottom-up input drives the membrane potential to 1 (governed by the multiplier of 1–v). The second source is the leak with the fixed term, L. The third source is the inhibition that results in the multiplication of inhibition strength, I, and a sum of two terms. One term is lateral inhibition which is the sum of output at time t from k units at the auditory layer. Another term is the inhibition from the motor layer, n^*m', which is specified next. The combination of the leak and inhibition drives the membrane potential toward 0 (as the term in the bracket is multiplied by –v). The updating speed is proportional to the integration rate (time constant, τ). The sum of three sources multiplying by the integration rate yields the updating magnitude for the membrane potential at each time. The fixed parameters are similar to those used in the previous study (Li et al., 2020).

The influences of motor signals are modeled as two sets of free parameters. The motor signals come from the same motor units but are split into two sources. One source integrates activities of all motor units into an interneuron that inhibits all auditory neurons (Fig. 5A). For simplification, the inhibition effect of each motor neuron is assigned as a unit value, m. The equivalent inhibition effects from the interneuron are the sum of n motor units, n^*m. This motor source simulates the hypothesized function of CD. Another source directly modulates corresponding auditory neurons. This motor signal is modeled as a gain control parameter, g_i, which increases the gain of excitatory input to the corresponding auditory unit. This motor source simulates the hypothesized function of EC.

The previous study (Li et al., 2020) combined the implementation of an interneuron and gain modulation in one neural network model to collaboratively simulate the inhibition and enhancement throughout the time course of speech production in normal participants. In this study, we manipulated these two key parameters of CD (the parameter of m) and SP (the parameter of g_i) to simulate the results of AVHs and non-AVHs groups. Specifically, during the simulation of GP, because of the hypothesis of CD impairment, we fitted the model to the observed suppression during GP by adjusting the parameter of inhibition strength, m', separately for AVHs and non-AVHs group (the apostrophe after the label of the parameter indicates changes from the value obtained from the normal group).

During the simulation of SP, only the prepared syllable in the motor layer is activated in normal participants. We hypothesized that the non-AVHs group would have intact EC function as normal participants, but the AVHs group would have a 'noisy' EC function. We fitted the model to the observed reversal of SP effects in AVHs by adjusting the parameter of g'_i. Specifically, for simulation of the AVHs group, the parameter of gain modulation, g'_i, was modified proportionally according to the prepared vs. unprepared syllables -- the gain from the prepared motor unit was decreased by X times, at the same time the gain from the unprepared motor units were increased by X times. That is, compared with normal participants and the non-AVHs group, the signal-to-noise ratio (i.e. the ratio between the gain to the prepared vs.

unprepared auditory nodes) was adjusted.

To assess how the model fitted the empirical results, we treated the model simulation results as the mean from a distribution with unknown variance. Data in GP and SP conditions were subject to one-sample *t*-tests against the simulation results, separately for non-AVHs and AVHs groups. Because this analysis was to test the null hypothesis that the model simulation results were from a similar distribution of empirical results, we used a Bayesian analysis method for one-sample *t*-tests (Rouder et al., 2009) (online tool at http://rdrr.io/rforge/BayesFactor/man/ttest.tstat.html). The Bayes factor is B01 = M0/M1, where M0 and M1 are the marginal likelihood for the null and alternative, respectively. That is, the Bayes factor is an odds ratio between the null and alternative hypotheses, which indicates that the null is B01 times more probable than the alternative. The parameters for the Bayesian analysis were a sample size of 19 and 16 for AVHs and non-AVHs groups and a scale *r* on an effect size of 0.707.

References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin, 141*(5), 931.

Bansal, S., Bae, G.-Y., Robinson, B.M., Hahn, B., Waltz, J., Erickson, M., . . . Gold, J.M. (2022). Association between failures in perceptual updating and the severity of psychosis in schizophrenia. *JAMA Psychiatry, 79*(2), 169-177.

Blakemore, S.J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2*(8), 561-567.

Blakemore, S.J., Wolpert, D.M., & Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience, 1*(7), 635-640.

Blanchard, J.J., & Neale, J.M. (1994). The neuropsychological signature of schizophrenia: generalized or differential deficit? *American Journal of Psychiatry*.

Carandini, M. (2012). From circuits to behavior: a bridge too far? *Nature Neuroscience, 15*(4), 507-509.

Chu, Q., Ma, O., Hang, Y., & Tian, X. (2022). Dual-stream cortical feedbacks mediate sensory prediction. *bioRxiv*.

Crapse, T.B., & Sommer, M.A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience, 9*(8), 587-600.

Deecke, L., Scheid, P., & Kornhuber, H.H. (1969). Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental Brain Research, 7*(2), 158-168.

Eliades, S.J., & Wang, X. (2005). Dynamics of auditory–vocal interaction in monkey auditory cortex. *Cerebral Cortex, 15*(10), 1510-1523.

Eliades, S.J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature, 453*(7198), 1102-1106.

Enikolopov, A.G., Abbott, L.F., & Sawtell, N.B. (2018). Internally generated predictions enhance neural and behavioral detection of sensory stimuli in an electric fish. *Neuron, 99*(1), 135-146.

Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophrenia Bulletin, 4*(4), 636.

Feinberg, I., & Guazzelli, M. (1999). Schizophrenia – a disorder of the corollary discharge systems that integrate the motor systems of thought with the sensory systems of consciousness. *British Journal of Psychiatry, 174*(3), 196-204. doi:10.1192/bjp.174.3.196

Fletcher, P.C., & Frith, C.D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience, 10*(1), 48-58.

Flinker, A., Chang, E.F., Kirsch, H.E., Barbaro, N.M., Crone, N.E., & Knight, R.T. (2010). Single-trial speech suppression of auditory cortex activity in humans.

Journal of Neuroscience, 30(49), 16643-16650.

Ford, J.M., & Mathalon, D.H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research, 38*(1), 37-46.

Ford, J.M., Mathalon, D.H., Heinks, T., Kalba, S., Faustman, W.O., & Roth, W.T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry, 158*(12), 2069-2071.

Ford, J.M., Mathalon, D.H., Kalba, S., Whitfield, S., Faustman, W.O., & Roth, W.T. (2001). Cortical responsiveness during talking and listening in schizophrenia: an event-related brain potential study. *Biological Psychiatry, 50*(7), 540-549.

Ford, J.M., Palzes, V.A., Roach, B.J., & Mathalon, D.H. (2014). Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia Bulletin, 40*(4), 804-812.

Ford, J.M., Roach, B.J., Faustman, W.O., & Mathalon, D.H. (2007). Synch before you speak: auditory hallucinations in schizophrenia. *American Journal of Psychiatry, 164*(3), 458-466.

Ford, J.M., Roach, B.J., Faustman, W.O., & Mathalon, D.H. (2008). Out-of-synch and out-of-sorts: dysfunction of motor-sensory communication in schizophrenia. *Biological Psychiatry, 63*(8), 736-743.

Friston, K.J., Stephan, K.E., Montague, R., & Dolan, R.J. (2014). Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry, 1*(2), 148-158.

Frith, C.D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine, 17*(3), 631-648.

Frith, C.D. (2015). *The Cognitive Neuropsychology of Schizophrenia: Classic Edition*: Psychology press.

Gold, J.M., Corlett, P.R., Strauss, G.P., Schiffman, J., Ellman, L.M., Walker, E.F., . . . Silverstein, S.M. (2020). Enhancing psychosis risk prediction through computational cognitive neuroscience. *Schizophrenia Bulletin, 46*(6), 1346-1352.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., . . . Hämäläinen, M.S. (2014). MNE software for processing MEG and EEG data. *Neuroimage, 86*, 446-460.

Heinks-Maldonado, T.H., Mathalon, D.H., Houde, J.F., Gray, M., Faustman, W.O., & Ford, J.M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of General Psychiatry, 64*(3), 286-296.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402. doi:10.1038/nrn2113

Hoffman, R.E., Hawkins, K.A., Gueorguieva, R., Boutros, N.N., Rachid, F., Carroll, K., & Krystal, J.H. (2003). Transcranial magnetic stimulation of left temporoparietal cortex and medication-resistant auditory hallucinations. *Archives of General Psychiatry, 60*(1), 49-56.

Houde, J.F., & Nagarajan, S.S. (2011). Speech production as state feedback control.

Frontiers in Human Neuroscience, 5, 82.

Houde, J.F., Nagarajan, S.S., Sekihara, K., & Merzenich, M.M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience, 14*(8), 1125-1138.

Huys, Q.J.M., Maia, T.V., & Frank, M.J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience, 19*(3), 404-413.

Jas, M., Engemann, D.A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *Neuroimage, 159*, 417-429. doi:https://doi.org/10.1016/j.neuroimage.2017.06.030

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology, 9*(6), 718-727.

Kay, S.R., Fiszbein, A., & Opler, L.A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin, 13*(2), 261-276.

Kilteni, K., & Ehrsson, H.H. (2017). Body ownership determines the attenuation of self-generated tactile sensations. *Proceedings of the National Academy of Sciences, 114*(31), 8426-8431.

Kosslyn, S.M., Ganis, G., & Thompson, W.L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2*(9), 635-642.

Li, S., Zhu, H., & Tian, X. (2020). Corollary discharge versus efference copy: distinct neural signals in speech preparation differentially modulate auditory responses. *Cerebral Cortex, 30*(11), 5806-5820.

Li, Y., Luo, H., & Tian, X. (2020). Mental operations in rhythm: Motor-to-sensory transformation mediates imagined singing. *PLoS Biology, 18*(10), e3000504.

Marcus, G., Marblestone, A., & Dean, T. (2014). The atoms of neural computation. *Science, 346*(6209), 551-552.

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*: MIT press.

McGuire, P.K., David, A.S., Murray, R.M., Frackowiak, R.S.J., Frith, C.D., Wright, I., & Silbersweig, D.A. (1995). Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. *Lancet, 346*(8975), 596-600.

Miall, R.C., & Wolpert, D.M. (1996). Forward models for physiological motor control. *Neural Networks, 9*(8), 1265-1279.

Montague, P.R., Dolan, R.J., Friston, K.J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences, 16*(1), 72-80.

Pinheiro, A.P., Schwartze, M., Gutiérrez-Domínguez, F., & Kotz, S.A. (2020). Real and imagined sensory feedback have comparable effects on action anticipation. *Cortex, 130*, 290-301.

Ross, J., Morrone, M.C., Goldberg, M.E., & Burr, D.C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24*(2), 113-121.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237.

Shergill, S.S., Brammer, M.J., Williams, S.C.R., Murray, R.M., & McGuire, P.K. (2000).

Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatry, 57*(11), 1033-1038.

Singla, S., Dempsey, C., Warren, R., Enikolopov, A.G., & Sawtell, N.B. (2017). A cerebellum-like circuit in the auditory system cancels responses to self-generated sounds. *Nature Neuroscience, 20*(7), 943-950.

Sperry, R.W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology, 43*(6), 482.

Stephan, K.E., Friston, K.J., & Frith, C.D. (2009). Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bulletin, 35*(3), 509-527.

Taschereau-Dumouchel, V., Michel, M., Lau, H., Hofmann, S.G., & LeDoux, J.E. (2022). Putting the "mental" back in "mental disorders": a perspective from research on fear and anxiety. *Molecular Psychiatry, 27*(3), 1322-1330.

Tian, X., Ding, N., Teng, X., Bai, F., & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour, 2*(3), 225-234.

Tian, X., & Huber, D.E. (2008). Measures of spatial similarity and response magnitude in MEG and scalp EEG. *Brain Topography, 20*(3), 131-141.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience, 6*, 314.

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience, 25*(7), 1020-1036.

Tian, X., Poeppel, D., & Huber, D.E. (2011). TopoToolbox: Using Sensor Topography to Calculate Psychologically Meaningful Measures from Event-Related EEG/MEG. *Computational Intelligence and Neuroscience, 2011*, 674605. doi:10.1155/2011/674605

Tian, X., Zarate, J.M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex, 77*, 1-12. doi:10.1016/j.cortex.2016.01.002

Toh, W.L., Moseley, P., & Fernyhough, C. (2022). Hearing voices as a feature of typical and psychopathological experience. *Nature Reviews Psychology, 1*(2), 72-86.

Ungerleider, L.G., & Haxby, J.V. (1994). 'What'and 'where'in the human brain. *Current Opinion in Neurobiology, 4*(2), 157-165.

von Holst, E., & Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften, 37*(20), 464-476.

Wang, X.-J., & Krystal, J.H. (2014). Computational psychiatry. *Neuron, 84*(3), 638-654.

Wang, X., Zhu, H., & Tian, X. (2019). Revealing the temporal dynamics in non-invasive electrophysiological recordings with topography-based analyses. *bioRxiv*, 779546.

Waters, F., Allen, P., Aleman, A., Fernyhough, C., Woodward, T.S., Badcock, J.C., . . . Menon, M. (2012). Auditory hallucinations in schizophrenia and nonschizophrenia populations: a review and integrated model of cognitive

mechanisms. *Schizophrenia Bulletin, 38*(4), 683-693.

Wheeler, M.E., Petersen, S.E., & Buckner, R.L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, 97*(20), 11125-11129.

Whitford, T.J. (2019). Speaking-induced suppression of the auditory cortex in humans and its relevance to schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4*(9), 791-804.

Wolpert, D.M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3*(11), 1212-1217.

Yang, F., Fang, X., Tang, W., Hui, L., Chen, Y., Zhang, C., & Tian, X. (2019). Effects and potential mechanisms of transcranial direct current stimulation (tDCS) on auditory hallucinations: A meta-analysis. *Psychiatry Research, 273*, 343-349.

Yang, F., Zhu, H., Yu, L., Lu, W., Zhang, C., & Tian, X. (2021). Deficits in multi-scale top-down processes distorting auditory perception in schizophrenia. *Behavioural Brain Research, 412*, 113411.

Yang, J., Zhu, H., & Tian, X. (2018). Group-level multivariate analysis in EasyEEG toolbox: examining the temporal dynamics using topographic responses. *Frontiers in Neuroscience, 12*, 468.

Zapparoli, L., Seghezzi, S., Zirone, E., Guidali, G., Tettamanti, M., Banfi, G., . . . Paulesu, E. (2020). How the effects of actions become our own. *Science Advances, 6*(27), eaay8301.

Zatorre, R.J., & Halpern, A.R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron, 47*(1), 9-12.

Zheng, X., Zhu, H., Li, S., & Tian, X. (2022). The generic inhibitory function of corollary discharge in motor intention: evidence from the modulation effects of speech preparation on the late components of auditory neural responses. *eNeuro*, 0309-0322.2022. doi:10.1523/ENEURO.0309-22.2022

Acknowledgements

This study was supported by the National Natural Science Foundation of China 32071099 and 32271101, Natural Science Foundation of Shanghai 20ZR1472100, Program of Introducing Talents of Discipline to Universities, Base B16018 to X.T., and East China Normal University (ECNU) Academic Innovation Promotion Program for Excellent Doctoral Students YBNLTS2019-026 to F.Y.

Author contributions

FY designed the experiment, recruited and evaluated schizophrenia patients, collected and analyzed data, interpreted the data, and drafted and edited the manuscript. HZ helped code the experimental methods and scripts. As psychiatrists, CZ , XF, LY and ZW helped recruit and evaluate schizophrenia patients. SL shared the normal control data. XT designed the experiment and edited the manuscript. All authors approved the final version.

Data and code availability

The analytical code and datasets are available from the corresponding author upon reasonable request.

Declaration of Competing Interest

The authors report no declarations of interest.